SENTIMENT ANALYSIS IN TWITTER

A Project Report

Submitted in the partial fulfillment of the requirement

for the award of the degree of

BACHELOR OF ENGINEERING

IN

INFORMATION TECHNOLOGY

BY

RAJARSHI SARKAR (BE/1397/2011)

AMIT KUMAR (BE/1513/2011)

UNDER GUIDANCE OF

DR. VIJAY KUMAR JHA

ASSOCIATE PROFESSOR

DEPT. OF COMPUTER SCIENCE AND ENGINEERING

BIRLA INSTITUTE OF TECHNOLOGY

MESRA, RANCHI – 835215

DECLARATION CERTIFICATE

This is to certify that the work presented in the project report entitled "**Sentiment Analysis in Twitter"**in partial fulfillment of the requirement for the award of Degree of Bachelor of Engineering in Information Technology of Birla Institute of Technology, Mesra, Ranchi is an authentic work carried out under my supervision and guidance.

To the best of my knowledge, the content of this project report does not form a basis for the award of any previous degree to anyone else.

Date:

Dr. Vijay Kumar Jha
Dept. of Computer Science & Engineering
Birla Institute of Technology
Mesra, Ranchi-835215

CERTIFICATE OF APPROVAL

The foregoing project entitled "**Sentiment Analysis in Twitter**", is hereby approved as a creditable study of research topic and has been presented in satisfactory manner to warrant its acceptance as prerequisite to the degree for which it has been submitted.

It is understood that by this approval, the undersigned do not necessarily endorse any conclusion drawn or opinion expressed therein, but approve the project report for the purpose for which it is submitted.

(Internal Examiner) **(External Examiner)**

(Dr. Sandip Dutta)

Head of the Department

Dept. of Computer Science & Engineering,

Birla Institute of Technology,

Mesra, Ranchi-835215

ACKNOWLEDGEMENT

We owe our deepest gratitude to our guide **Dr. Vijay Kumar Jha**, who helped us throughout the project. He made sure that we learnt by practice, always motivating us to think a step further, work a little harder. Our interactions with him always resulted in newer ideas and proved beneficial towards our work. Without his constant presence and supervision our work would not have been successful.

We especially acknowledge the many useful discussions we had among ourselves that helped us understand some of the subtle technical problems in a better way.

Rajarshi Sarkar

(BE/1397/2011)

Amit Kumar

(BE/1513/2011)

ABSTRACT

Twitter is a popular microblogging service where users create status messages (called "tweets"). These tweets sometimes express opinions about different topics. Sentiment Analysis in Twitter is a method to automatically extract sentiment (positive or neutral or negative) from a tweet. This is very useful because it allows feedback to be aggregated without manual intervention. Consumers can use sentiment analysis to research products or services before making a purchase. Marketers can use this to research public opinion of their company and products, or to analyze customer satisfaction. Organizations can also use this to gather critical feedback about problems in newly released products. There has been a large amount of research in the area of sentiment classification. Traditionally most of it has focused on classifying larger pieces of text, like reviews. Tweets (and microblogs in general) are different from reviews primarily because of their purpose: while reviews represent summarized thoughts of authors in a specific topic, tweets are short messages about a variety of topics and are limited to 140 characters of text. Also, the frequency of misspellings, acronyms and slang in tweets is much higher than in other domains.

TABLE OF CONTENTS

1. INTRODUCTION

Twitter is a popular microblogging service where users create status messages (called "tweets"). These tweets sometimes express opinions about different topics. We propose a method to automatically extract sentiment (positive or neutral or negative) from a tweet. This is very useful because it allows feedback to be aggregated without manual intervention. Consumers can use sentiment analysis to research products or services before making a purchase. Marketers can use this to research public opinion of their company and products, or to analyze customer satisfaction. Organizations can also use this to gather critical feedback about problems in newly released products. There has been a large amount of research in the area of sentiment classification. Traditionally most of it has focused on classifying larger pieces of text, like reviews. Tweets (and microblogs in general) are different from reviews primarily because of their purpose: while reviews represent summarized thoughts of authors, tweets are more casual and limited to 140 characters of text. Generally, tweets are not as thoughtfully composed as reviews. Yet, they still offer companies an additional avenue to gather feedback. Previous research on analyzing blog posts includes. Pang et al. have analyzed the performance of different classifiers on movie reviews. The work of Pang et al. has served as a baseline and many authors have used the techniques provided in their paper across different domains. In order to train a classifier, supervised learning usually requires hand-labeled training data. With the large range of topics discussed on Twitter, it would be very difficult to manually collect enough data to train a sentiment classifier for tweets. Hence, we have used publicly available twitter datasets which are in turn obtained via distant supervision proposed. However, this dataset consist only of positive and negative tweets. For neutral tweets, we have used the publicly available neutral tweet dataset provided. We run the machine learning classifiers Naïve Bayes ,Support Vector Machine trained on the positive and negative tweets dataset and the neutral tweets against a test set of tweets. To help visualize the utility of the Twitter-based sentiment analysis tool, we have built a web application tool. This can be used by individuals and companies that may want to research sentiment on any topic.

1.1 Defining the sentiment

For the purpose of research, we define sentiment to be “a personal positive or negative feeling” and when there is an absence of this, we treat it as a neutral sentiment.Table 1 shows some examples.

Table 1. Example Tweets

Sentiment	**Keyword**	**Tweet**
Positive	Football	Dammmmm we Love Football
Neutral	airplane	Comes 8 a clock, phone going on airplane mode
Negative	Pep Guardiola	Pep Guardiola to resign as Barcelona boss

1.2 Characteristic of Tweets

Twitter messages have many unique attributes, which differentiates my work from previous research:

- **Length :-**The maximum length of a Twitter message is 140 characters. This is very different from the previous sentiment classification research that focused on classifying longer bodies of work, such as movie reviews.
- **Language model :-**Twitter users post messages from many different media, including their cell phones. The frequency ofmisspellings and slang in tweets is much higher than in other domains.
- **Domain :-**Twitter users post short messages about a variety of topics unlike other sites which are tailored to a specific topic. This differs from a large percentage of past research, which focused on specific domains such as movie reviews.

2. WHY SENTIMENT ANALYSIS IN TWITTER? [Ref: 1,2]

Twitter is a popular micro blogging service where users create status messages (called "tweets"). These tweets sometimes express opinions about different topics. We propose to build an automatic sentiment (positive or neutral or negative) extractor from a tweet. This is very useful because it allows feedback to be aggregated without manual intervention. Using this analyzer,

- Consumers can use sentiment analysis to research products or services before making a purchase. E.g. Kindle
- Marketers can use this to research public opinion of their company and products, or to analyze customer satisfaction. E.g. Election Polls
- Organizations can also use this to gather critical feedback about problems in newly released products. E.g. Brand Management (Nike, Adidas)

3. OBJECTIVE AND SCOPE OF THE PROJECT

The objective is to be able to automatically classify a tweet as apositive tweet or a neutral tweet or a negative tweet based on itssentiment.
Using this analyzer

- Consumers can use sentiment analysis to research products or services before making a purchase [Ref: 3]. E.g. Kindle
- Marketers can use this to research public opinion of their company and products, or to analyze customer satisfaction [Ref: 4]. E.g. Election Polls
- Organizations can also use this to gather critical feedback about problems in newly released products. E.g. Brand Management (Mufti,Puma).

4. TOOLS USED

- **Python**[Ref: 5]**:-**Python is a widely used general-purpose, high-level programming language. Its design philosophy emphasizes codereadability, and its syntax allows programmers to express concepts in fewer lines of code than would be possible in languages such as C++ or Java.

  The language provides constructs intended to enable clear programs on both a small and large scale.Python supports multiple programming paradigms, including object-oriented, imperative and functional programming or proceduralstyles. It features a dynamic type system and automatic memory management and has a large and comprehensive standard library.Python interpreters are available for installation on many operating systems, allowing Python code execution on a wide variety of systems.

  Using third-party tools, such as Py2exe or Pyinstaller, Python code can be packaged into stand-alone executable programs for some of the most popular operating systems, allowing for the distribution of Python-based software for use on those environments without requiring the installation of a Python interpreter.

- **Natural Language Toolkit (NLTK)**[Ref: 6]**:-** The Natural Language Toolkit, or more commonly NLTK, is a suite of libraries and programs for symbolic and statistical natural language processing (NLP) for the Python programming language. NLTK includes graphical demonstrations and sample data.

  It is accompanied by a book that explains the underlying concepts behind the language processing tasks supported by the toolkit, plus a cookbookis intended to support research and teaching in NLP or closely related areas, including empirical linguistics, cognitive science,artificial intelligence, information retrieval, and machine learning. NLTK has been used successfully as a teaching tool, as an individual study tool, and as a platform for prototyping and building research systems.

- **LIBSVM**[Ref: 7]**:-** LIBSVM is popular open source machine learning library, developed at the National Taiwan University and written in C++ though with a C API. LIBSVM implements the SMO algorithm for kernelizedsupport vector machines (SVMs), supporting classification and regression.

5. FETCHING TWEETS

To access the live stream, you will need to install the oauth2 library so you can properly authenticate.This library is already installed on the class virtual machine, but you can install it yourself in your Python environment. (The command $ pip install oauth2 should work for most environments.)
The steps below will help you set up your twitter account to be able to access the live 1% stream.
1.Create a twitter account if you do not already have one.

2.Go to **https://dev.twitter.com/apps** and log in with your twitter credentials.

3.Click “Create New App”

4.Fill out the form and agree to the terms. Put in a dummy website if you don’t have one you want touse.

5. On the next page, click the “API Keys” tab along the top, then scroll all the way down until you seethe section “Your Access Token”

6. Click the button “Create My Access Token”. You can Read more about Oauth authorization.

7. You will now copy four values into twitterstream.py. These values are your “API Key”, your “APIsecret”, your “Access token” and your “Access token secret”. All four should now be visible on theAPI Keys page. (You may see “API Key” referred to as “Consumer key” in some places in the code oron the web; they are synonyms.)
Open twitterstream.py and set the variables corresponding to the apikey, api secret, access token, and access secret. You will see code like the below:
api_key = “<Enter api key>”
api_secret = “<Enter api secret>”
access_token_key = “<Enter your access token key here>”
access_token_secret = “<Enter your access token secret here>”

8. Run the following and make sure you see data flowing and that no errors occur.
$ python twitterstream.py > tweets.txt
This command pipes the output to a file. Stop the program with CtrlC,
but wait at least 3 minutes for data to accumulate. Keep the file tweets.txt for the duration of the assignment; we will be reusing it in later problems. Don’t use someone else’s file; we will check for uniqueness in other parts of the assignment.

9. If you wish, modify the file to use the twitter search API to search for specific terms. For example, tosearch for the term “5abeller5”, you can pass the following url to the twitterreq function:
https://api.twitter.com/1.1/search/tweets.json?q=microsoft

6. PREPROCESSING TWEETS

Preprocessing tweets includes:

- Getting the tweet from the “text” _eld of a JSON document.
- Convert the tweets to lower case.
- Eliminate all of these URLs via regular expression matching or replace with generic word URL.
- Eliminate ‘@username’ via regex matching or replace it with generic word AT USER.
- Hashtags can give us some useful information, so it is useful to replace them with the exact same word without the hash.

E.g.#nike replaced with ‘nike’.

- Remove punctuation at the start and ending of the tweets..

E.g:‘the day is beautiful!’ replaced with ‘the day is beautiful’.

- Remove additional white space.

CODE:-

```
#import regex

import re

#start process_tweet

def processTweet(tweet):

# process the tweets

#Convert to lower case

tweet = tweet.lower()

#Convert www.* or https?://* to URL

tweet = re.sub(‘((www\.[^\s]+)|(https?://[^\s]+))’,’URL’,tweet)

#Convert @username to AT_USER

tweet = re.sub(‘@[^\s]+’,’AT_USER’,tweet)
```

```
#Remove additional white spaces
tweet = re.sub(‘[\s]+’, ‘‘, tweet)
#Replace #word with word
tweet = re.sub(r’#([^\s]+)’, r’\1’, tweet)
#trim
tweet = tweet.strip(‘\’”’)
return tweet
#end
#Read the tweets one by one and process it
fp = open(‘data/sampleTweets.txt’, ‘r’)
line = fp.readline()
while line:
processedTweet = processTweet(line)
print processedTweet
line = fp.readline()
#end loop
fp.close()
```

7. APPROACHES

Sentiment Dictionary Approach:-A file having 2482 words and their sentiment value (ranging from -5 to 5) is maintained. We call this file Sentiment Dictionary.
A word and its sentiment in the Sentiment Dictionary is seperated by a tab character, i.e., '\t'. A very positive word like 'happy' is given a sentiment value of 5. A neutral word like 'aeroplane' is given a sentiment value close to 0. A very negative word like 'hate' is given a sentiment value of -5.

Naive Bayes Approach:-In machine learning, naive Bayes classifiers are a family of simple probabilistic classifiers based on applying Bayes' theorem with strong (naive) independence assumptions between the features. Naive Bayes classifiers are highly scalable, requiring a number of parameters linear in the number of variables (features/predictors) in a learning problem. Maximum-likelihood training can be done by evaluating a closed-form expression, which takes linear time, rather than by expensive iterative approximation as used for many other types of classifiers In the statistics and computer science literature, Naive Bayes models are known under a variety of names, including simple Bayesand independence Bayes. All these names reference the use of Bayes' theorem in the classifier's decision rule, but naive Bayes is not (necessarily) a Bayesian method;Russell and Norvig note that "[naive Bayes] is sometimes called a Bayesian classifier, a somewhat careless usage that has prompted true Bayesians to call it the idiot Bayes model.

Support Vector Machines Approach:- In machine learning, support vector machines (SVMs, also support vector networks) are supervised learning models with associated learning algorithms that analyze data and recognize patterns, used for classification and regression analysis. Given a set of training examples, each marked as belonging to one of two categories, an SVM training algorithm builds a model that assigns new examples into one category or the other, making it a non-probabilistic binary linear classifier. An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall on.

In addition to performing linear classification, SVMs can efficiently perform a non-linear classification using what is called the kernel trick, implicitly mapping their inputs into high-dimensional feature spaces.

7.1. SENTIMENT DICTIONARY APPROACH

7.1.1. Overview[Ref: 8]:-

A file having 2482 words and their sentiment value (ranging from -5 to 5) is maintained. We call this file Sentiment Dictionary.
A word and its sentiment in the Sentiment Dictionary is 9abeller9 by a tab character, i.e., '\t'. A very positive word like 'happy' is given a sentiment value of 5. A neutral word like 'aeroplane' is given a sentiment value close to 0. A very negative word like 'hate' is given a sentiment value of -5.

7.1.2. Algorithm[Appendix:A]:-

The algorithm for Sentiment Dictionary Approach is as follows:

- Get all the words present in the tweet.
- Add the sentiment value of all words while referring to the Sentiment Dictionary. If the word is not present in the Sentiment Dictionary then add 0.
- If the cumulative sentiment value is positive then the tweet is positive.
- If the cumulative sentiment value is zero then the tweet is neutral.
- If the cumulative sentiment value is negative then the tweet is negative.

EXAMPLE:-

Words in tweet	We	Like	Cars
Sentiment value	0	2	0

Table: Cumulative sentiment value: 2. Tweet is classified as positive.

Words in tweet	We	have	cars
Sentiment value	0	0	0

Table: Cumulative sentiment value: 0. Tweet is classified as neutral.

Words in tweet	We	hate	cars
Sentiment value	0	-3	0

Table: Cumulative sentiment value: -3. Tweet is classified as negative.

7.1.3. Difficulties Faced

Problems in Sentiment Dictionary Approach are as follows:

- If all the the words of the tweet are not present in the Sentiment Dictionary then the tweet will be judged as neutral which is not necessarily true always.
- If a tweet has positive and negative words as per the Sentiment Dictionary, then they can very well add upto 0. Thereby, the tweet will be judged as neutral which is not necessarily true always.

7.1.4. Results

1000 testing tweets (367 positive, 418 neutral and 215 negative tweets) were tested against 2482 words of the Sentiment Dictionary.

The results are tabulated below:

	Judged as		
	Positive	**Neutral**	**Negative**
Positive Testing Tweets (367)	206	152	9
Neutral Testing Tweets (418)	33	358	27
Negative Testing Tweets (215)	17	59	139

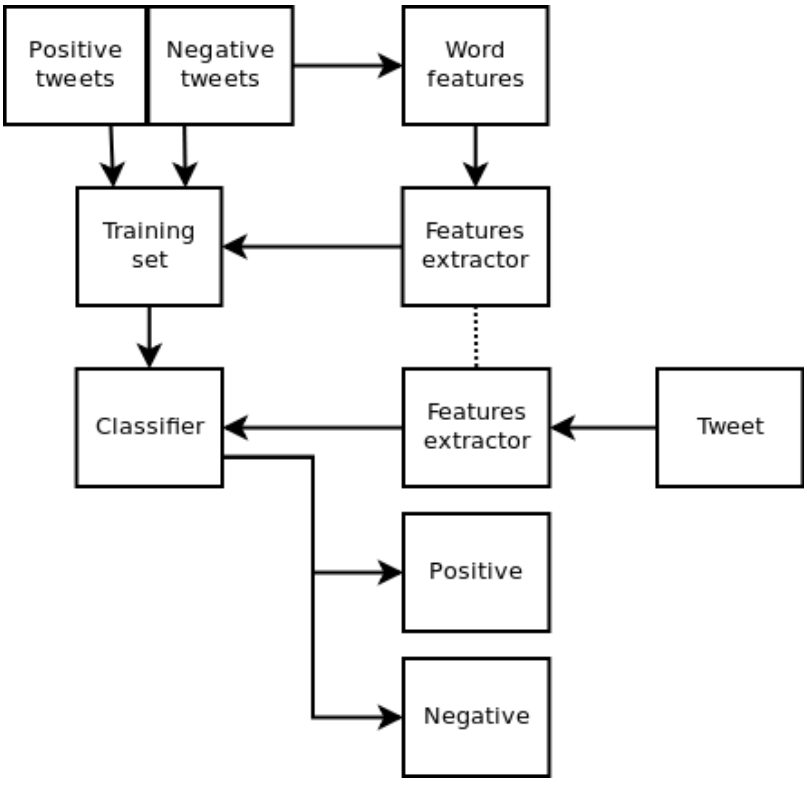
Table: Sentiment Dictionary Approach Analysis.

Overall Accuracy: 70.3 %

7.2. NAIVE BAYES APPROACH

7.2.1. Overview[Ref: 9,10]:-

In machine learning, naive Bayes classifiers are a family of simple probabilistic classifiers based on applying Bayes' theorem with strong (naive) independence assumptions between the features. Naive Bayes classifiers are highly scalable, requiring a number of parameters linear in the number of variables (features/predictors) in a learning problem. Maximum-likelihood training can be doneby evaluating a closed-form expression, which takes linear time, rather than by expensive iterative approximation as used for many other types of classifiers In the statistics and computer science literature, Naive Bayes models are known under a variety of names, including simple Bayesand independence Bayes. All these names reference the use of Bayes' theorem in the classifier's decision rule, but naive Bayes is not (necessarily) a Bayesian method;Russell and Norvig note that "[naive Bayes] is sometimes called a Bayesian classifier, a somewhat careless usage that has prompted true Bayesians to call it the idiot Bayes model.



Naive Bayes Classifier

7.2.2. Algorithm[Appendix:B]:-

- Preprocess the testing tweets.
- Remove all the stopwords (this, we, etc.) from the training and testing set.
- Estimate the probability P(c) of each class c by dividing the number of words in tweets in c by the total number of words in the training data set.
- Estimate the probability distribution P(w | c) for all words w and classes c. This can be done by dividing the number of occurences of w in tweets in c by the total number of words in c.
- To find the score of a tweet t for class c, calculate:

$$\text{score}(t, c) = P(c) * \prod_{i=1}^{n} P(wi / c)$$

- To predict the most likely class label, just pick the c with the highest score value.

EXAMPLE:-

There are 16905 words (5830, 5320, 5755 words in the positive, neutral and negative respectively) in the training data set.

Words in tweet	We	Like	cars
Probability(positive)	-	8/5830	1/5830
Probability(neutral)	-	2/5320	3/5320
Probability(negative)	-	1/5755	1/5755

Positive score	5830/16905 * 8/5830 * 1/5830 = 0.000000081
Neutral score	5320/16905 * 2/5320 * 3/5320 = 0.000000067
Negative score	5755/16905 * 1/5755 * 1/5755 = 0.000000010

Tweet classified as: Positive.

7.2.3. Difficulties Faced

Problems in Naive Bayes Approach are as follows:

- It assumes each feature to be independent of all other features. This is the "naive" assumption seen in the multiplication of P(wi /c) in the de_nition of score. Thus, for example, if there is a feature 'best' and another 'world's best', then their probabilities would be multiplied as though independent, even though the two are overlapping.

- The same issues arise for words that are highly correlated with other words (idioms, phrases, etc.).

7.2.4. Results

1000 testing tweets (367 positive, 418 neutral and 215 negative tweets) were tested against 1500 tweets (500 positive, 500 neutral and 500 negative tweets). The results are tabulated below:

	Judged as		
	Positive	**Neutral**	**Negative**
Positive Testing Tweets (367)	194	89	84
Neutral Testing Tweets (418)	24	347	47
Negative Testing Tweets (215)	21	53	141

Table: Naive Bayes Approach Analysis.

Overall Accuracy: 68.2 %

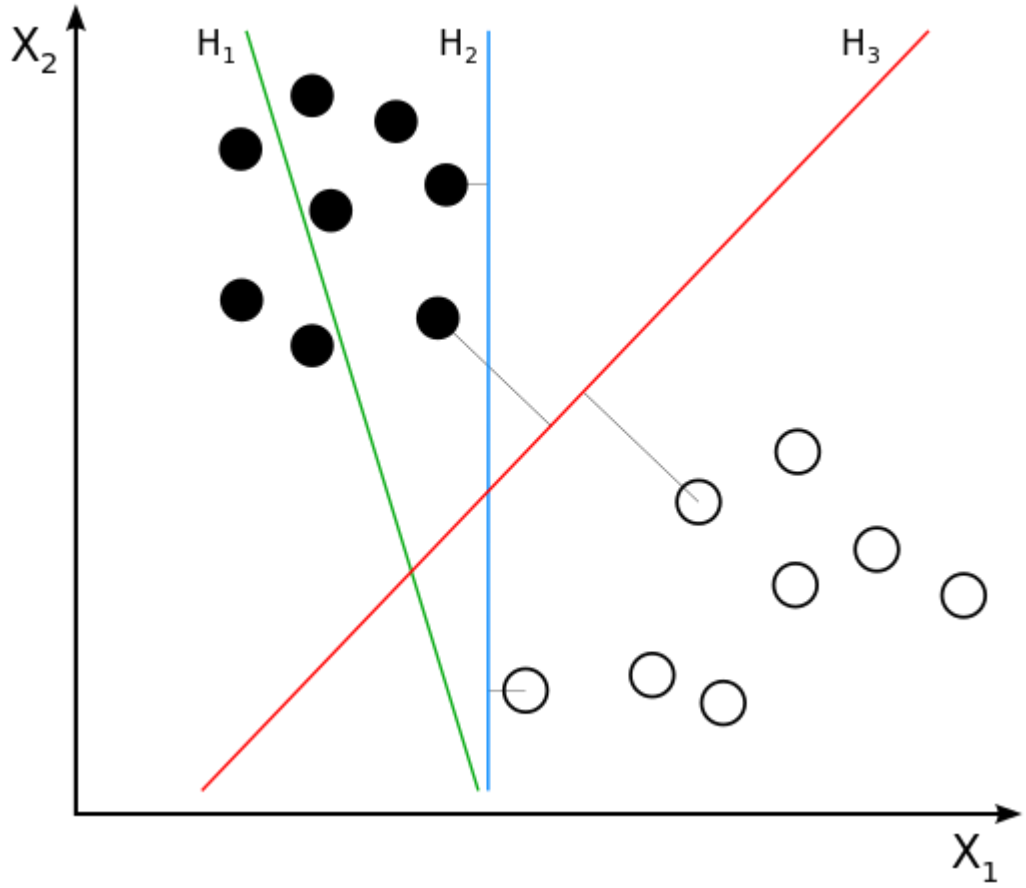
7.3. SUPPORT VECTOR MACHINES APPROACH

7.3.1. Overview [Ref: 12]:-

In machine learning, support vector machines (SVMs, also support vector networks) are supervised learning models with associated learning algorithms that analyze data and recognize patterns, used for classification and regression analysis. Given a set of training examples, each marked as belonging to one of two categories, an SVM training algorithm builds a model that assigns new examples into one category or the other, making it a non-probabilistic binary linear classifier. An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall on.

In addition to performing linear classification, SVMs can efficiently perform a non-linear classification using what is called the kernel trick, implicitly mapping their inputs into high-dimensional feature spaces.

- Supervised learning model with associated learning algorithms that analyzes data and is used for classification.
- A data point is viewed as a p dimensional vector, and we want to know whether we can separate such points with a (p-1) dimensional hyperplane.
- The best hyperplane is the one that represents the largest separation, or margin, between the two classes.
- New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall on.



H1 does not separate the classes. H2 does, but only with a small margin. H3 separates them with the maximum margin.

- Given some training data D, a set of n points of the form:

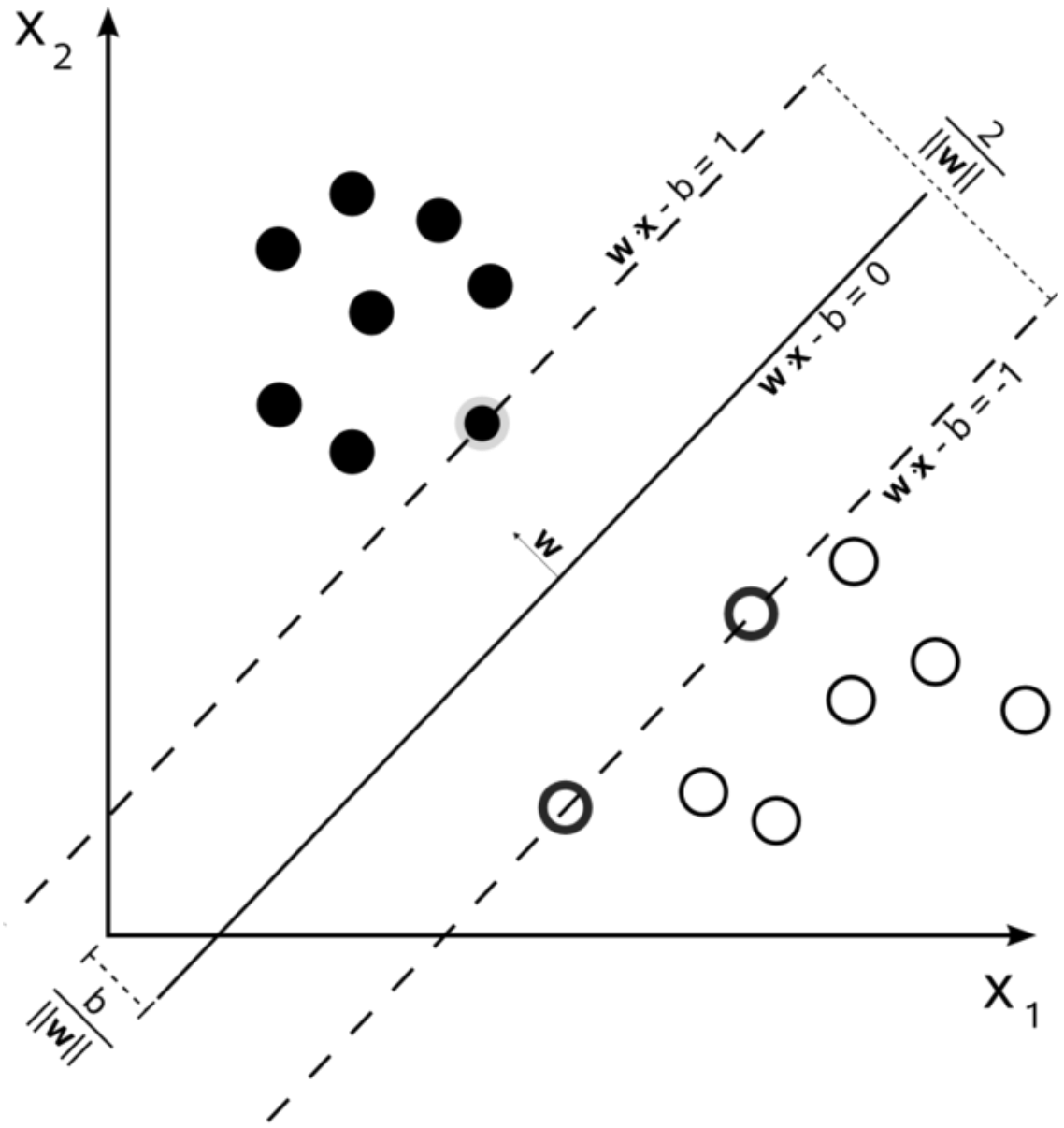
$$D = \{ (x_i ; y_i) \mid x_i \ \epsilon y_i \in \{-1; 1\}$$

- Any hyperplane can be written as the set of points x satisfying:

$$w.x - b = 0:$$

where . denotes the dot product and w the normal vector to the hyperplane.

- By using geometry, we find the distance between these two hyperplanes is $2/\|w\|$, so we want to minimize $\|w\|$.



Maximum-margin hyperplane and margins for an SVM trained withsamples from two classes. Samples on the margin are called thesupport vectors.

7.3.2. Algorithm [Appendix:C]:-

Algorithm implemented using LIBSVM is shown below:

- Train the linear multiclass SVM classifier based on training tweet data set . Each training tweet is mapped into the hyperplanewith help of its feature list.
- Preprocess the testing tweets and build a feature list for eachtweet.
- Map each testing tweet into the hyperplane with help of itsfeature list to know the class of the tweet.

7.3.3. Difficulties Faced

Problems in Support Vector Machines Approach are as follows:

- Parameters of a solved model were 18abeller18 to interpret.
- Length of feature list is huge.

7.3.4. Results

1000 testing tweets (367 positive, 418 neutral and 215 negative tweets) were tested against 4550 tweets (2441 positive, 689 neutral and 1720 negative tweets). The results are tabulated below:

	Judged as		
	Positive	**Neutral**	**Negative**
Positive Testing Tweets (367)	272	52	42
Neutral Testing Tweets (418)	73	332	13
Negative Testing Tweets (215)	23	39	153

Table: Support Vector Machines Approach Analysis.

Overall Accuracy: 75.7 %

8.FUTURE WORK

Machine learning techniques perform well for classifying sentiment in tweets. We believe the accuracy of the system could be still improved. Below is a list of ideas we think could help the classification:-

Maximum entropy classifier[Ref: 13] The idea behind Maximum Entropy models is that one should prefer the most uniform models that satisfy a given constraint . MaxEnt models are feature-based models. In a two class scenario, it is the same as using logistic regression to find a distribution over the classes. MaxEnt makes no independence assumptions for its features, unlike Naive Bayes.

Random-forest: It is an ensemble-learning method for classification, regression and other tasks, that operate by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees. Random forests correct for decision trees' habit of overfitting to their training set. [Ref: 14]

Semantics [Ref: 15]The algorithms classify the overall sentiment of a tweet. The polarity of a tweet may depend on the perspective you are interpreting the tweet from. For example, in the tweet "Federer beats Nadal ☺", the sentiment is positive for Federer and negative for Nadal. In this case, semantics may help. Using a semantic role 20abeller may indicate which noun is mainly associated with the verb and the classification would take place accordingly. This may allow "Nadal beats Federer ☺" to be classified differently from "Federer beats Nadal ☺".

Bigger Dataset[Ref: 16]The training dataset in the order of millions will cover a better range of twitter words and hence better unigram feature vector resulting in an overall improved model. This would vastly improve upon the existing classifier results.

InternationalizationCurrently, we focus only on English tweets but Twitter has a huge international audience. It should be possible to use my approach to classify sentiment in other languages with a language specific positive/negative keyword list.

9.CONCLUSION

Using a novel feature vector of weighted unigrams, we have shown that machine learning algorithms such as Naïve Bayes and Support Vector Machines achieve competitive accuracy in classifying tweet sentiment.

Following is the summary of the accuracies of all the approaches used:

Approach used	**Accuracy**
Sentiment Dictionary	70.3 %
Naive Bayes	68.2 %
Support Vector Machines	75.7 %

Also, accuracy of classification depends on the dataset and the method used to classify.

Appendix A: Sentiment Dictionary Approach Codes

twitterstream.py

```
# python twitterstream.py > tweets.txt


import oauth2 as oauth

import urllib2 as urllib


access_token_key = "288801977-
8tPnj2yukfZedG9AD281VM8W6Ny58etWRHeNUSfN"

access_token_secret =
"cpSCLTOWcWOQypy1r9qy1ckwhSaEpEiRvm4WJCEwrvgnt"

consumer_key = "dACKtQhEYVSUmrt9amAehPZTS"

consumer_secret =
"jeM514YDEuI0EppABKKl0WWYlGTtKtrxLaEs9jY0vojPw7Dzah"

oauth_token    = oauth.Token(key=access_token_key, secret=access_token_secret)

oauth_consumer = oauth.Consumer(key=consumer_key, secret=consumer_secret)

signature_method_hmac_sha1 = oauth.SignatureMethod_HMAC_SHA1()

http_method = "GET"

http_handler  = urllib.HTTPHandler(debuglevel=0)

https_handler = urllib.HTTPSHandler(debuglevel=0)


def twitterreq(url, method, parameters): # Construct, sign, and open a twitter
request using the credentials above.
```

```
  req = oauth.Request.from_consumer_and_token(oauth_consumer,
token=oauth_token, http_method=http_method, http_url=url, parameters=parameters)


  req.sign_request(signature_method_hmac_sha1, oauth_consumer, oauth_token)
  headers = req.to_header()
  if http_method == "POST":
    encoded_post_data = req.to_postdata()
  else:
    encoded_post_data = None
    url = req.to_url()

  opener = urllib.OpenerDirector()
  opener.add_handler(http_handler)
  opener.add_handler(https_handler)
  response = opener.open(url, encoded_post_data)

  return response

def fetchsamples():
  url = "https://stream.twitter.com/1/statuses/sample.json"
  parameters = []
  response = twitterreq(url, "GET", parameters)
  for line in response:
    print line.strip()

if __name__ == '__main__':
```

fetchsamples()

tweet_sentiment.py

```
# python tweet_sentiment.py Sentiment-Dictionary.txt tweets.txt > output.txt

# python tweet_sentiment.py Sentiment-Dictionary.txt tweets.txt


import sys

import json

scores = {}  # initialize an empty dictionary


def dictionary(file):

    global scores

    sentimentdictionaryfile = file

    for line in sentimentdictionaryfile:

        (term, score) = line.split('\t')# The file is tab-delimited. "\t" means "tab character"

        scores[term] = int(score)# Convert the score to an integer.


def tweets(file):

    i = 0

    for line in file.readlines():

        tweet = json.loads(line)

        tweet_score = 0

        if 'text' in tweet and 'lang' in tweet and tweet['lang'] == 'en':
```

```
        spectweet = tweet['text'].encode('utf-8')
        for word in spectweet.lower().split():
            if word in scores.keys():
                tweet_score += scores[word]
        i = i + 1
        print 'Tweet No.',
        print i,
        print ':',
        spectweet = spectweet.replace('\n', '.')
        spectweet = spectweet.replace('\r', '.')
        print spectweet
        print 'Sentiment',
        print i,
        print ':',
        if float(tweet_score) > 0:
                        print "We think that the sentiment was positive in that
sentence."
        if float(tweet_score) < 0:
                        print "We think that the sentiment was negative in that
sentence."
        if float(tweet_score) == 0:
                        print "We think that the sentiment was neutral in that
sentence."
        #print float(tweet_score)


def main():
    sent_file = open(sys.argv[1])
```

```
    tweet_file = open(sys.argv[2])
    dictionary(sent_file)
    tweets(tweet_file)

if __name__ == '__main__':
    main()
```

find_new_terms.py

```
# python find_new_terms.py Sentiment-Dictionary.txt tweets.txt
# script that computes the sentiment for the terms that do not appear in the file Sentiment-Dictionary.txt.

import sys
import json
scores = {} # initialize an empty dictionary

def dictionary(file):
    global scores
    sentimentdictionaryfile = file
    for line in sentimentdictionaryfile:
        term, score  = line.split("\t")  # The file is tab-delimited. "\t" means "tab character"
        scores[term] = int(score)  # Convert the score to an integer.
```

```
def tweets(file):
    for line in file.readlines():
        tweet = json.loads(line)
        if 'text' in tweet and ('lang' in tweet and tweet['lang']=="en"):
            spectweet = tweet['text'].encode('utf-8')
            for word in spectweet.lower().split():
                if word not in scores.keys():
                    print word

def main():
    sent_file = open(sys.argv[1])
    tweet_file = open(sys.argv[2])
    dictionary(sent_file)
    tweets(tweet_file)

if __name__ == '__main__':
    main()
```

run.py

```
# python run.py > output.txt

import os
os.system('python twitterstream.py > tweets.txt')
os.system('python tweet_sentiment.py Sentiment-Dictionary.txt tweets.txt')
```

Appendix B: Naive Bayes Approach Codes

NaiveBayes_TrainingTweets.py

```
# python NaiveBayes_TrainingTweets.py tweets.txt > output.txt


import nltk
import math
import re
import sys
import os
import codecs
import json
reload(sys)
sys.setdefaultencoding('utf-8')
from nltk.corpus import stopwords


#Pull out all of the words in a list of tagged tweets, formatted in tuples.
def getwords(tweets):
    allwords = []
    for (words, sentiment) in tweets:
        allwords.extend(words)
    return allwords
```

```
#Order a list of tweets by their frequency.
def getwordfeatures(listoftweets):
#Print out wordfreq if you want to have a look at the individual counts of words.
        wordfreq = nltk.FreqDist(listoftweets)
        words = wordfreq.keys()
        return words

    def feature_extractor(doc):
        docwords = set(doc)
        features = {}
        for i in wordlist:
            features['contains(%s)' % i] = (i in docwords)
        return features

    customstopwords =
['a','about','above','across','after','again','against','all','almost','alone','along','already','als
o','although','always','among','an','and','another','any','anybody','anyone','anything','any
where','are','area','areas','around','as','ask','asked','asking','asks','at','away','b','back','back
ed','backing','backs','be','became','because','become','becomes','been','before','began','be
hind','being','beings','best','better','between','big','both','but','by','c','came','can','cannot','
case','cases','certain','certainly','clear','clearly','come','could','d','did','differ','different','di
fferently','do','does','done','down','downed','downing','downs','during','e','each','early','ei
ther','end','ended','ending','ends','enough','even','evenly','ever','every','everybody','every
one','everything','everywhere','f','face','faces','fact','facts','far','felt','few','find','finds','firs
t','for','four','from','full','fully','further','furthered','furthering','furthers','g','gave','general'
,'generally','get','gets','give','given','gives','go','going','goods','got','greater','greatest','gro
up','grouped','grouping','groups','h','had','has','have','having','he','her','here','herself','hig
h','higher','highest','him','himself','his','how','however','i','if','important','in','interest','int
erested','interesting','interests','into','is','it','its','itself','j','just','k','keep','keeps','kind','kne
w','know','known','knows','l','large','largely','last','later','latest','least','less','let','lets','like'
,'likely','long','longer','longest','m','made','make','making','man','many','may','me','mem
```

ber','members','men','might','more','most','mostly','mr','mrs','much','must','my','myself','not','n','necessary','need','needed','needing','needs','never','new','newer','newest','next','no','nobody','non','noone','now','nowhere','number','numbers','o','of','off','often','old','older','oldest','on','once','one','only','open','opened','opening','opens','or','order','ordered','ordering','orders','other','others','our','out','over','p','part','parted','parting','parts','per','perhaps','place','places','point','pointed','pointing','points','possible','present','presented','presenting','presents','problem','problems','put','puts','q','quite','r','rather','really','right','room','rooms','s','said','same','saw','say','says','second','seconds','see','seem','seemed','seeming','seems','sees','several','shall','she','should','show','showed','showing','shows','side','sides','since','small','smaller','smallest','so','some','somebody','someone','something','somewhere','state','states','still','such','sure','t','take','taken','than','that','the','their','them','then','there','therefore','these','they','thing','things','think','thinks','this','those','though','thought','thoughts','three','through','thus','to','today','together','too','took','toward','turn','turned','turning','turns','two','u','under','until','up','upon','us','use','used','uses','v','very','w','want','wanted','wanting','wants','was','way','ways','we','well','wells','went','were','what','when','where','whether','which','while','who','whole','whose','why','will','with','within','without','work','worked','working','works','would','x','y','year','years','yet','you','young','younger','youngest','your','yours','z']

taggedtweets = [("the rock is destined to be the 21st century's new "" conan "" and that he's going to make a splash even greater than arnold schwarzenegger , jean-claud van damme or steven segal .","positive"),

("the gorgeously elaborate continuation of "" the lord of the rings "" trilogy is so huge that a column of words cannot adequately describe co-writer/director peter jackson's expanded vision of j . r . r . tolkien's middle-earth .","positive"),

("effective but too-tepid biopic","positive"),

("if you sometimes like to go to the movies to have fun , wasabi is a good place to start .","positive"),

("emerges as something rare , an issue movie that's so honest and keenly observed that it doesn't feel like one .","positive"),

("the film provides some great insight into the neurotic mindset of all comics -- even those who have reached the absolute top of the game .","positive"),

("offers that rare combination of entertainment and education .","positive"),

("perhaps no picture ever made has more literally showed that the road to hell is paved with good intentions .","positive"),

("steers turns in a snappy screenplay that curls at the edges ; it's so clever you want to hate it . but he somehow pulls it off .","positive"),

("take care of my cat offers a refreshingly different slice of asian cinema .","positive"),

("this is a film well worth seeing , talking and singing heads and all . makes me smile .","positive"),

("what really surprises about wisegirls is its low-key quality and genuine tenderness .","positive"),

("(wendigo is) why we go to the cinema : to be fed through the eye , the heart , the mind .","positive"),

("one of the greatest family-oriented , fantasy-adventure and awesome movies ever .","positive"),

("ultimately , it ponders the reasons we need stories so much .","positive"),

("an utterly compelling 'who wrote it' in which the reputation of the most famous author who ever lived comes into question .","positive"),

("illuminating if overly talky documentary .","positive"),

("a masterpiece four years in the making .","positive"),

("the movie's ripe , enrapturing beauty will tempt those willing to probe its inscrutable mysteries .","positive"),

("offers a breath of the fresh air of true sophistication .","positive"),

("a thoughtful , provocative , good, great and insistently humanizing film .","positive"),

("with a cast that includes some of the top actors working in independent film , lovely & amazing involves us because it is so incisive , so bleakly amusing about how we go about our lives .","positive"),

("a disturbing and frighteningly evocative assembly of imagery and hypnotic music composed by philip glass .","positive"),

("not for everyone , but for those with whom it will connect , it's a nice departure from standard moviegoing fare .","positive"),

("scores a few points for doing what it does with a dedicated and good-hearted professionalism .","positive"),

("occasionally melodramatic , it's also extremely effective .","positive"),

("spiderman rocks and makes everyone smile .","positive"),

("an idealistic love story that brings out the latent 15-year-old romantic in everyone .","positive"),

("at about 95 minutes , treasure planet maintains a brisk pace as it races through the familiar story . however , it lacks grandeur and that epic quality often associated with stevenson's tale as well as with earlier disney efforts .","positive"),

("it helps that lil bow wow . . . tones down his pint-sized gangsta act to play someone who resembles a real kid .","positive"),

("guaranteed to move anyone who ever shook , rattled , or rolled .","positive"),

("a masterful film from a master filmmaker , unique in its deceptive grimness , compelling in its fatalist worldview .","positive"),

("light , cute and forgettable .","positive"),

("if there's a way to effectively teach kids about the dangers of drugs , i think it's in projects like the (unfortunately r-rated) paid .","positive"),

("while it would be easy to give crush the new title of two weddings and a funeral , it's a far more thoughtful film than any slice of hugh grant whimsy .love in the air.","positive"),

("though everything might be literate and smart , it never took off and always seemed static .","positive"),

("cantet perfectly captures the hotel lobbies , two-lane highways , and roadside cafes that permeate vincent's days","positive"),

("ms . fulford-wierzbicki is almost spooky in her sulky , calculating lolita turn .","positive"),

("though it is by no means his best work , laissez-passer is a distinguished and distinctive effort by a bona-fide master , a fascinating film replete with rewards to be had by all willing to make the effort to reap them .","positive"),

("like most bond outings in recent years , some of the stunts are so outlandish that they border on being cartoonlike . a heavy reliance on cgi technology is beginning to creep into the series .","positive"),

("newton draws our attention like a magnet , and acts circles around her better known co-star , mark wahlberg .","positive"),

("the story loses its bite in a last-minute happy ending that's even less plausible than the rest of the picture . much of the way , though , this is a refreshingly novel ride full of love.","positive"),

("fuller would surely have called this gutsy and at times exhilarating movie a great yarn .","positive"),

("the film makes a strong case for the importance of the musicians in creating the motown sound .","positive"),

("karmen moves like rhythm itself , her lips chanting to the beat , her long , braided hair doing little to wipe away the jeweled beads of sweat .","positive"),

("gosling provides an amazing performance that dwarfs everything else in the film .","positive"),

("a real movie , about real people , that gives us a rare glimpse into a culture most of us don't know . love it.","positive"),

("tender yet lacerating and darkly funny fable .","positive"),

("may be spoofing an easy target -- those old '50's giant creature features -- but . . . it acknowledges and celebrates their cheesiness as the reason why people get a kick out of watching them today .","positive"),

("an engaging overview of johnson's eccentric career .","positive"),

("in its ragged , cheap and unassuming way , the movie works .","positive"),

("some actors have so much charisma that you'd be happy to listen to them reading the phone book . hugh grant and sandra bullock are two such likeable actors .","positive"),

("sandra nettelbeck beautifully orchestrates the transformation of the chilly , neurotic , and self-absorbed martha as her heart begins to open . smile . :)","positive"),

("behind the snow games and lovable siberian huskies (plus one sheep dog) , the picture hosts a parka-wrapped dose of heart .","positive"),

("everytime you think undercover brother has run out of steam , it finds a new way to surprise and amuse .","positive"),

("manages to be original , even though it rips off many of its ideas .","positive"),

("you'd think by now america would have had enough of plucky british eccentrics with hearts of gold . yet the act is still charming here .","positive"),

("whether or not you're enlightened by any of derrida's lectures on "" the other "" and "" the self , "" derrida is an undeniably fascinating and playful fellow .","positive"),

("a pleasant enough movie , held together by skilled ensemble actors . full of love.","positive"),

("this is the best american movie about troubled teens since 1998's whatever . makes me smile . :)","positive"),

("disney has always been hit-or-miss when bringing beloved kids' books to the screen . . . tuck everlasting is a little of both .","positive"),

("just the labour involved in creating the layered richness of the imagery in this chiaroscuro of madness and light is astonishing .","positive"),

("the animated subplot keenly depicts the inner struggles of our adolescent heroes - insecure , uncontrolled , and intense .","positive"),

("the invincible werner herzog is alive and well and living in la","positive"),

("morton is a great actress portraying a complex character , but morvern callar grows less compelling the farther it meanders from its shocking start .","positive"),

("part of the charm of satin rouge is that it avoids the obvious with humour and lightness .","positive"),

("son of the bride may be a good half-hour too long but comes replete with a flattering sense of mystery and quietness .","positive"),

("a simmering psychological drama in which the bursts of sudden violence are all the more startling for the slow buildup that has preceded them .","positive"),

("a taut , intelligent psychological drama .","positive"),

("a compelling coming-of-age drama about the arduous journey of a sensitive young girl through a series of foster homes and a fierce struggle to pull free from her dangerous and domineering mother's hold over her .","positive"),

("a truly moving experience , and a perfect example of how art -- when done right -- can help heal , clarify , and comfort .","positive"),

("this delicately observed story , deeply felt and masterfully stylized , is a triumph for its maverick director .","positive"),

("at heart the movie is a deftly wrought suspense yarn whose richer shadings work as coloring rather than substance .","positive"),

("the appearance of treebeard and gollum's expanded role will either have you loving what you're seeing , or rolling your eyes . i loved it ! gollum's 'performance' is incredible !","positive"),

("a screenplay more ingeniously constructed than "" memento ","positive"),

("if this movie were a book , it would be a page-turner , you can't wait to see what happens next .","positive"),

("haneke challenges us to confront the reality of sexual aberration .","positive"),

("absorbing and disturbing -- perhaps more disturbing than originally intended -- but a little clarity would have gone a long way .","positive"),

("it's the best film of the year so far , the benchmark against which all other best picture contenders should be measured .","positive"),

("painful to watch , but viewers willing to take a chance will be rewarded with two of the year's most accomplished and riveting film performances .","positive"),

("this is a startling film that gives you a fascinating , albeit depressing view of iranian rural life close to the iraqi border .","positive"),

("an imaginative comedy/thriller . leaves you with a smile .","positive"),

("a few artsy flourishes aside , narc is as gritty as a movie gets these days .","positive"),

("while the isle is both preposterous and thoroughly misogynistic , its vistas are incredibly beautiful to look at .","positive"),

("together , tok and o orchestrate a buoyant , darkly funny dance of death . in the process , they demonstrate that there's still a lot of life in hong kong cinema .","positive"),

("director kapur is a filmmaker with a real flair for epic landscapes and adventure , and this is a better film than his earlier english-language movie , the overpraised elizabeth .","positive"),

("the movie is a blast of educational energy , as bouncy animation and catchy songs escort you through the entire 85 minutes .","positive"),

("a sports movie with action that's exciting on the field and a story you care about off it .","positive"),

("doug liman , the director of bourne , directs the traffic well , gets a nice wintry look from his locations , absorbs us with the movie's spycraft and uses damon's ability to be focused and sincere .","positive"),

("the tenderness of the piece is still intact .","positive"),

("katz uses archival footage , horrifying documents of lynchings , still photographs and charming old reel-to-reel recordings of meeropol entertaining his children to create his song history , but most powerful of all is the song itself","positive"),

("like the film's almost anthropologically detailed realization of early-'80s suburbia , it's significant without being overstated .","positive"),

("while mcfarlane's animation lifts the film firmly above the level of other coming-of-age films . . . it's also so jarring that it's hard to get back into the boys' story .","positive"),

("if nothing else , this movie introduces a promising , unusual kind of psychological horror .","positive"),

("writer-director burger imaginatively fans the embers of a dormant national grief and curiosity that has calcified into chronic cynicism and fear .","positive"),

(". . . a roller-coaster ride of a movie","positive"),

("i enjoyed time of favor while i was watching it , but i was surprised at how quickly it faded from my memory . and this makes me smile .","positive"),

("chicago is sophisticated , brash , sardonic , completely joyful in its execution .","positive"),

("steve irwin's method is ernest hemmingway at accelerated speed and volume .","positive"),

("a refreshing korean film about five female high school friends who face an uphill battle when they try to take their relationships into deeper waters .","positive"),

("while the mystery unravels , the characters respond by hitting on each other .","negative"),

("britney spears' phoniness is nothing compared to the movie's contrived , lame screenplay and listless direction .","negative"),

("every sequel you skip will be two hours gained . consider this review life-affirming .","negative"),

("if the movie were all comedy , it might work better . but it has an ambition to say something about its subjects , but not a willingness .","negative"),

("the movie , while beautiful , feels labored , with a hint of the writing exercise about it .","negative"),

("twenty-three movies into a mostly magnificent directorial career , clint eastwood's efficiently minimalist style finally has failed him . big time .","negative"),

("this heist flick about young brooklyn hoods is off the shelf after two years to capitalize on the popularity of vin diesel , seth green and barry pepper . it should have stayed there .","negative"),

("the film has a childlike quality about it . but the feelings evoked in the film are lukewarm and quick to pass .","negative"),

("the most opaque , self-indulgent and just plain goofy an excuse for a movie as you can imagine .","negative"),

("it's not a film to be taken literally on any level , but its focus always appears questionable .","negative"),

("big fat liar is little more than home alone raised to a new , self-deprecating level .","negative"),

("the movie is gorgeously made , but it is also somewhat shallow and art-conscious .","negative"),

("the only time 8 crazy nights comes close to hitting a comedic or satirical target is during the offbeat musical numbers .","negative"),

("loses its sense of humor in a vat of failed jokes , twitchy acting , and general boorishness .","negative"),

("there's a delightfully quirky movie to be made from curling , but brooms isn't it .","negative"),

("the story suffers a severe case of oversimplification , superficiality and silliness .","negative"),

("chamber of secrets will find millions of eager fans . but if the essence of magic is its make-believe promise of life that soars above the material realm , this is the opposite of a truly magical movie .","negative"),

("too clever by about nine-tenths .","negative"),

("has all the hallmarks of a movie designed strictly for children's home video , a market so insatiable it absorbs all manner of lame entertainment , as long as 3-year-olds find it diverting .","negative"),

("well-meant but unoriginal .","negative"),

("bears about as much resemblance to the experiences of most battered women as spider-man does to the experiences of most teenagers .","negative"),

("toward the end sum of all fears morphs into a mundane '70s disaster flick .","negative"),

("director carl franklin , so crisp and economical in one false move , bogs down in genre cliches here .","negative"),

("mendes still doesn't quite know how to fill a frame . like the hanks character , he's a slow study : the action is stilted and the tabloid energy embalmed .","negative"),

("this thing is just garbage .","negative"),

("as crimes go , writer-director michael kalesniko's how to kill your neighbor's dog is slight but unendurable .","negative"),

("there must be an audience that enjoys the friday series , but i wouldn't be interested in knowing any of them personally .","negative"),

("a bold (and lovely) experiment that will almost certainly bore most audiences into their own brightly colored dreams .","negative"),

("an uplifting , largely bogus story .","negative"),

("an empty exercise , a florid but ultimately vapid crime melodrama with lots of surface flash but little emotional resonance .","negative"),

("if you are curious to see the darker side of what's going on with young tv actors (dawson leery did what ? ! ?) , or see some interesting storytelling devices , you might want to check it out , but there's nothing very attractive about this movie .","negative"),

("my own minority report is that it stinks .","negative"),

("trying to make head or tail of the story in the hip-hop indie snipes is enough to give you brain strain -- and the pay-off is negligible .","negative"),

("the script is high on squaddie banter , low on shocks .","negative"),

(". . . if you , like me , think an action film disguised as a war tribute is disgusting to begin with , then you're in for a painful ride .","negative"),

("while solondz tries and tries hard , storytelling fails to provide much more insight than the inside column of a torn book jacket .","negative"),

("with very little to add beyond the dark visions already relayed by superb recent predecessors like swimming with sharks and the player , this latest skewering . . . may put off insiders and outsiders alike .","negative"),

("[davis] wants to cause his audience an epiphany , yet he refuses to give us real situations and characters .","negative"),

("without a fresh infusion of creativity , 4ever is neither a promise nor a threat so much as wishful thinking .","negative"),

(". . . unlike [scorsese's mean streets] , ash wednesday is essentially devoid of interesting characters or even a halfway intriguing plot .","negative"),

("being unique doesn't necessarily equate to being good , no matter how admirably the filmmakers have gone for broke .","negative"),

("a few hours after you've seen it , you forget you've been to the movies .","negative"),

("odd and weird .","negative"),

("waydowntown may not be an important movie , or even a good one , but it provides a nice change of mindless pace in collision with the hot oscar season currently underway .","negative"),

("yes , i suppose it's lovely that cal works out his issues with his dad and comes to terms with his picture-perfect life -- but world traveler gave me no reason to care , so i didn't .","negative"),

("shadyac , who belongs with the damned for perpetrating patch adams , trots out every ghost trick from the sixth sense to the mothman prophecies .","negative"),

("the photographer's show-don't-tell stance is admirable , but it can make him a problematic documentary subject .","negative"),

("it is not the first time that director sara sugarman stoops to having characters drop their pants for laughs and not the last time she fails to provoke them .","negative"),

("i'd be hard pressed to think of a film more cloyingly sappy than evelyn this year .","negative"),

("nothing more than an amiable but unfocused bagatelle that plays like a loosely-connected string of acting-workshop exercises .","negative"),

("meanders between its powerful moments .","negative"),

("what remains is a variant of the nincompoop benigni persona , here a more annoying , though less angry version of the irresponsible sandlerian manchild , undercut by the voice of the star of road trip .","negative"),

("a backhanded ode to female camaraderie penned by a man who has little clue about either the nature of women or of friendship .","negative"),

("pure of intention and passably diverting , his secret life is light , innocuous and unremarkable .","negative"),

(". . . delivers few moments of inspiration amid the bland animation and simplistic story .","negative"),

("take away the controversy , and it's not much more watchable than a mexican soap opera .","negative"),

("it's got the brawn , but not the brains . bad","negative"),

("mindless and boring martial arts and gunplay with too little excitement and zero compelling storyline .","negative"),

("a lot of talent is wasted in this crass , low-wattage endeavor .","negative"),

("to show these characters in the act and give them no feelings of remorse -- and to cut repeatedly to the flashback of the original rape -- is overkill to the highest degree .","negative"),

("[t]oo many of these gross out scenes . very bad . .","negative"),

("about one in three gags in white's intermittently wise script hits its mark ; the rest are padding unashamedly appropriated from the teen-exploitation playbook .","negative"),

("little is done to support the premise other than fling gags at it to see which ones shtick .","negative"),

("reno does what he can in a thankless situation , the film ricochets from humor to violence and back again , and ryoko hirosue makes us wonder if she is always like that .","negative"),

("if jews were catholics , this would be catechism","negative"),

("one of those films that seems tailor made to air on pay cable to offer some modest amusements when one has nothing else to watch . It is very bad","negative"),

("the big ending surprise almost saves the movie . it's too bad that the rest isn't more compelling .","negative"),

("charming , if overly complicated . . .","negative"),

("schneider's mugging is relentless and his constant need to suddenly transpose himself into another character undermines the story's continuity and progression .","negative"),

("all very stylish and beautifully photographed , but far more trouble than it's worth , with fantasy mixing with reality and actors playing more than one role just to add to the confusion .","negative"),

("it's probably not easy to make such a worthless film . . .","negative"),

("hope keeps arising that the movie will live up to the apparent skills of its makers and the talents of its actors , but it doesn't .","negative"),

("has no reason to exist , other than to employ hollywood kids and people who owe favors to their famous parents .","negative"),

("for a guy who has waited three years with breathless anticipation for a new hal hartley movie to pore over , no such thing is a big letdown .","negative"),

("constantly slips from the grasp of its maker .","negative"),

("smothered by its own solemnity .","negative"),

("'christian bale's quinn [is] a leather clad grunge-pirate with a hairdo like gandalf in a wind-tunnel and a simply astounding cor-blimey-luv-a-duck cockney accent . '","negative"),

("might be one of those vanity projects in which a renowned filmmaker attempts to show off his talent by surrounding himself with untalented people .","negative"),

("after you laugh once (maybe twice) , you will have completely forgotten the movie by the time you get back to your car in the parking lot .","negative"),

("not one moment in the enterprise didn't make me want to lie down in a dark room with something cool to my brow .","negative"),

("in the era of the sopranos , it feels painfully redundant and inauthentic .","negative"),

("the overall vibe is druggy and self-indulgent , like a spring-break orgy for pretentious arts majors .","negative"),

("breen's script is sketchy with actorish notations on the margin of acting .","negative"),

("there's no question that epps scores once or twice , but it's telling that his funniest moment comes when he falls about ten feet onto his head .","negative"),

("if only merchant paid more attention the story .","negative"),

("at the one-hour mark , herzog simply runs out of ideas and the pace turns positively leaden as the movie sputters to its inevitable tragic conclusion .","negative"),

(". . . too contrived to be as naturally charming as it needs to be .","negative"),

("a simpler , leaner treatment would have been preferable ; after all , being about nothing is sometimes funnier than being about something .","negative"),

("the characters are based on stock clichs , and the attempt to complicate the story only defies credibility .","negative"),

("everything about it from the bland songs to the colorful but flat drawings is completely serviceable and quickly forgettable .","negative"),

("not the great american comedy , but if you liked the previous movies in the series , you'll have a good time with this one too .","negative"),

("a domestic melodrama with weak dialogue and biopic cliches .","negative"),

("mr . goyer's loose , unaccountable direction is technically sophisticated in the worst way .","negative"),

("the movie is so thoughtlessly assembled .","negative"),

("benigni presents himself as the boy puppet pinocchio , complete with receding hairline , weathered countenance and american breckin meyer's ridiculously inappropriate valley boy voice .","negative"),

("plays like some corny television production from a bygone era","negative"),

("the end result is like cold porridge with only the odd enjoyably chewy lump .","negative"),

("for all the charm of kevin kline and a story that puts old-fashioned values under the microscope , there's something creepy about this movie .","negative"),

("i was feeling this movie until it veered off too far into the exxon zone , and left me behind at the station looking for a return ticket to realism .","negative"),

("producer john penotti surveyed high school students . . . and came back with the astonishing revelation that "" they wanted to see something that didn't talk down to them . "" ignoring that , he made swimfan anyway","negative"),

("RT: Collobrate and present. New version of Google presentations. http://t.co/xmF34T27 VIA @googledownunder #google #googlepresentations","neutral"),

("Sdk available now, so you can develop apps for the phone and tablets right now. #Android4.0 #Google","neutral"),

("Great hardware and great new version of Android. we want it now. API is available now at least. #Google #Android #ICS #Development","neutral"),

("Hey #google ,when unveiling a new product,use a backdrop w/ lots of spatial-temporal high resolution activity. ;)","neutral"),

("Android beam, share any piece of information.from.one Android device to another by simply touching the devices. #Android4.0 #Google","neutral"),

("New Galaxy Nexus: App Improvements - Beam - sharing using NFC #nexus #samsung #google #android bit.ly/nEJbyE","neutral"),

("@jcmwright we are a going over to #Google right now, and we are in Maryland (#MICA). Catch up with us this week sometime #edu11","neutral"),

("Mashable! - Google Ice Cream Sandwich, Nexus Prime Launch [LIVE BLOG] #google #android #ice http://t.co/jtE7VuDK","neutral"),

("The new #google phone. Yoooo http://t.co/sr4sCat7","neutral"),

("New Galaxy Nexus: App Improvements - Quick Response to unwanted calls, predefined SMS's #nexus #samsung #google #android bit.ly/nEJbyE","neutral"),

("RT @jcmwright: Will hear from other small colleges who went #Google recently. Particularly interested in the learning tech prep ...","neutral"),

("Many new features for Android. #NexusPrime #Google","neutral"),

("#google must b having sum underlying reasoning 4 naming products after eatables #eclair #gingerbread #icecream #rawadosa #android","neutral"),

("Can they just tell us when #ICS will be available for Nexus S users? #Samsung #Google #HongKong","neutral"),

("RT @thedroidguy: #Google announces dates for 2012 IO http://t.co/90xoKyQN","neutral"),

("#Android #Google Google officially announces Ice Cream Sandwich http://t.co/Ij292Ye3 #DhilipSiva","neutral"),

("#Android #Google Samsung Galaxy Nexus launching in US, Europe and Asia this November http://t.co/NLdTU5oY #DhilipSiva","neutral"),

("#Google's open source search to end: http://t.co/h8PYttyr","neutral"),

("Google officially announces Ice Cream Sandwich http://t.co/u702sosO #android #google","neutral"),

("Samsung Galaxy Nexus launching in US, Europe and Asia this November http://t.co/SzchgKIf #android #google","neutral"),

("Hugo is buddies with 50 Cent? #google #android","neutral"),

("New Galaxy Nexus: App Improvements - New People App to improve contact information #nexus #samsung #google #android bit.ly/nEJbyE","neutral"),

("@5in_n_the_Air #GOOGLE time kml","neutral"),

("#Google announces dates for 2012 IO http://t.co/90xoKyQN","neutral"),

("we'm thinking about Google http://t.co/ACjRL4FO @GetGlue #Google","neutral"),

("Samsung Galaxy Nexus announced, full specs available http://t.co/VNGazg48 #samsung #galaxy #nexus #google #android #ics #thetechcheck","neutral"),

("Samsung Galaxy Nexus announced, full specs available http://t.co/0Nd3LiwV #samsung #galaxy #nexus #google #android #ics #thetechcheck","neutral"),

("#google #galaxy #nexus, we WANT A PHONE PLEASE TALK ABOUT OTHER FEATURES NOW THE CAMERA IS ONLY ONE PART!","neutral"),

("Funny how #samsung didn't have the #facebook app on their #galaxynexus demo but instead had #google+ LOOL.","neutral"),

("On Google+ then go here http://t.co/86xwWN2d #GPlus #Googleplus #Google #teamfollowback #socialnetwork","neutral"),

("New Galaxy Nexus: Hardware Improvements - Camera captures 1080p video #nexus #samsung #google #android bit.ly/nEJbyE","neutral"),

("Video of livestream of the Google/Samsung announcement froze. Damn my slow internet connection. #Google #Samsung","neutral"),

("1080p video recording, continuous focus, zoom while recording and time lapse included in Android 4.0. #Google #Android #ICS","neutral"),

("The new panorama feature in #Android Ice Cream Sandwich is pretty cool. Saw a glitch in the live presentation though. #Google","neutral"),

("Built in panorama in camera app #Android 4.0 #ICS #Google","neutral"),

("Google, Samsung unveil Galaxy Nexus phone running Android 4.0 http://t.co/C4elM79o"" via:@appleinsider #tech #google #Dubai #beirut","neutral"),

("Panorama","neutral"),

("#SamsungGalaxyNexus #google #android #IceCreamSandwich. New browser http://t.co/Zx6pIo6f","neutral"),

("now watching live stream of #Google new mobile #Nexus","neutral"),

("Damn group meeting disturbin me from watchin the #google event","neutral"),

("New tools giving the user the ability to restrict mobile data usage. #Google #Android #ICS","neutral"),

("Article discussing how the four great tech companies will compete in the marketplace: http://t.co/YFWF9iye #Apple #Google #Facebook #Amazon","neutral"),

("#Google releases an Infinite Digital Bookcase RT @VentureBeat http://t.co/YWtVgwJY by @MeghanKel","neutral"),

("Chrome Experiment - WebGL Bookcase - http://t.co/1GNsOwmu #Google","neutral"),

("Word of Mouth and the Internet - YouTube http://t.co/UCD9sDx4 #google #searching #wom","neutral"),

("@AtlantaSnoop yea we've seen my #location waaaay off on #google sites. Not a big deal but still strange. #spookygoogle","neutral"),

("#Android #Google Galaxy Nexus Site is Live: Register and Relive the New Features http://t.co/7llRSpUO #DhilipSiva","neutral"),

("On Google+ then go here http://t.co/iLjdQBpT #GPlus #Googleplus #Google #teamfollowback #socialnetwork","neutral"),

("#Google & Samsung announced their (hopeful) rebuttal to iPhone 4S tonight... site shows nice specs but no carriers yet. http://t.co/Z8SDgUT9","neutral"),

("RT @GWGoddess: We could have told you this two years ago! http://t.co/6BZ6W7cf @Novell #GroupWise #Google","neutral"),

("@Affan Fact! They still do everything better though #google","neutral"),

("@jpobrien11 or windows bridge for windows phones and don't try and pretend apple didn't steal the notification bar from #google #droidtweets","neutral"),

("What could a bookcase look like in 10 years...maybe this? http://t.co/i6YWQ7oR #google #books","neutral"),

("And thanks to TWIT's AAA crew for the live coverage of the #Google / Samsung announcement!","neutral"),

("Check%RT @CadientGroup: $6% of women use #Google for info on health care vs 28% of men -> http://t.co/QEfhPgX2 (via @nicolaziady) #hcsmeu","neutral"),

("Infinite digital dusting RT @VentureBeat: #Google releases an Infinite Digital Bookcase http://t.co/8m7gN6iN","neutral"),

("The competition: #Google introduces #Android 4.0 Ice Cream Sandwich and the Galaxy Nexus http://t.co/PCu88z7j","neutral"),

("Hide the women and children, break out the guns, #Google is going to encrypt your searches by default! http://t.co/2VbCGLgp #SEO","neutral"),

("#Android #Google Android 4.0 Ice Cream Sandwich SDK is available today http://t.co/mTnbj9hD #DhilipSiva","neutral"),

("RT @BrightSideNews: @Google and @Samsung Announce the #Galaxy #Nexus http://t.co/dZHvFxVh #Android #ICS #Icecreamsandwich #Google #Samsung","neutral"),

("So #Google #Droid is built (in part) using the bouncy castle C# api. C# is a Microsoft programming language. +1 for all you #Linux Fanboys","neutral"),

("My thoughts on tonight's #Google Ice Cream Sandwich and #Samsung Galaxy Nexus talk: http://t.co/gu7ScKXL","neutral"),

("#Galaxy #Nexus Officially Announced At Hong Kong Event http://t.co/wMy6LoCd #google","neutral"),

("#Android #Google Ice Cream Sandwich SDK now available http://t.co/K0tksZ1O #DhilipSiva","neutral"),

("Ice Cream Sandwich SDK now available http://t.co/ZOgECNER #android #google","neutral"),

("@csg122 http://t.co/TBxLrvin Hopefully you check twitter often enough to find this article. :D #google #ICS","neutral"),

("Why is everyone hating on #Android #IceCreamSandwich? #ICS #everyoneisacritic #os #google","neutral"),

("#Android #Google Galaxy Nexus Press Release http://t.co/JKrzYGBO #DhilipSiva","neutral"),

("engineers #google Manny Marroquin #MannyMarroquin","neutral"),

("@Google and @Samsung Announce the #Galaxy #Nexus http://t.co/dZHvFxVh #Android #ICS #Icecreamsandwich #Google #Samsung","neutral"),

("Will the #Galaxy #Nexus be coming to #Sprint? No mention of price, carriers, or source release for ICS. Give us those info #Google!","neutral"),

("RT @aalkhubaizi: Introducing Android 4.0, Ice Cream SandwichIntroducing Android 4.0, Ice Cream Sandwich http://t.co/L4Hqkv0c #Google #An ...","neutral"),

("#Advertising #Blog Today's blog covers the #IGen and #Google's ""Parisian"" Debating with @smarch323 for your votes! http://t.co/x3JY4b2L","neutral"),

("RT @DarkoIvancevic: How #Google Ventures Chooses Which Startups Get Its $200 Million http://t.co/9WDpnVDR","neutral"),

("#Google announces #NFC-based #Android Beam for sharing between phones http://t.co/zYKEr3ax via @engadget","neutral"),

("#Android #Google Android 4.0 Ice Cream Sandwich Official, SDK Now Available http://t.co/rXuPIW2U #DhilipSiva","neutral"),

("On Google+ then go here http://t.co/kDHaUfDl #GPlus #Googleplus #Google #teamfollowback #socialnetwork","neutral"),

("Samsung Galaxy Nexus - the new official Google smartphone with Android 4.0 ""Ice-Cream Sandwich"" http://t.co/S8hSL06q #Google #Android","neutral"),

("Installing the Ice Cream Sandwich SDK #android #google","neutral"),

("nothing like saying 'screw you' to a real estate company on 7 different social pages! #google #citysearch #yellowpages #yelp #yahoo ...","neutral"),

("Cool Infographic: Perks working for the big techs like #Google & #Facebook | http://t.co/lNEgf7Kn","neutral"),

("On Google+ then go here http://t.co/K6BggvDx #GPlus #Googleplus #Google #teamfollowback #socialnetwork","neutral"),

("#Google To Begin Encrypting Searches & Outbound Clicks By Default With SSL Search http://t.co/xW7vh75e","neutral"),

("#droidtweak #Video: #IceCreamSandwich on the #GalaxyNexus http://t.co/vBcl0OzP #googlephone #nexus1 #NexusS #google","neutral"),

("is oxycodone and nyquil a mix? lol .. we don't think so but.. #ineedadoctor and we'm to lazy to use #google","neutral"),

("Big Money... http://t.co/Ms6AwiX4 #seo #search #google #hack #sem #it #business #web #marketing #online #yes #hot #winning","neutral"),

("RT @RUILIFESTYLE: #GOOGLE THE MIXTAPE http://t.co/v8zTtNVV","neutral"),

("Massive Galaxy Nexus/Ice Cream Sandwich Recap. - http://t.co/I8SW5Udz #galaxynexus #Google #ics #Samsung","neutral"),

("#Android #Google Three UK to carry Samsung Galaxy Nexus http://t.co/Lt4m2HH4 #DhilipSiva","neutral"),

("#Android #Google Google updates Nexus site with Galaxy Nexus details http://t.co/BeckhmA1 #DhilipSiva","neutral"),

("Three UK to carry Samsung Galaxy Nexus http://t.co/SwvAcbxC #android #google","neutral"),

("Google updates Nexus site with Galaxy Nexus details http://t.co/LaSn4ol0 #android #google","neutral"),

("RT @tatn: #Google, #Samsung unveil Ice Cream Sandwich-powered Galaxy Nexus http://t.co/uY7KWJiY via @CNET #Android","neutral"),

("#google+ ...thoughts? to sign up or not to sign up, that is the question...","neutral"),

("#Google Ice Cream Sandwich, Nexus Prime Launch [LIVE BLOG] http://t.co/YD4FBog3 #uncategorized #android","neutral"),

("#Google has been released #Android 4.0 platform http://t.co/mvdCD0v8","neutral"),

("#Google Defaults to Encrypted HTTPS #Searches for Logged In Users [#Security] http://t.co/kzMZDxmE","neutral"),

("#Android #Google Samsung's Galaxy Nexus - It's Official, It's Headed For Global Availability, And It's... http://t.co/TI9Hy8LW #DhilipSiva","neutral"),

("Hopefully only a few weeks until #ICS is on my Nexus S 4G! #Google #Samsung #IceCreamSandwich #Android4","neutral"),

("#google just invented #mango taste #icecream. even the #roboto is so #segoe. but it's ok. every oem will have to pay #ms anyway... LOL.","neutral")]

```
tweets = []

#Create a list of words in the tweet, within a tuple.
for (word, sentiment) in taggedtweets:
    word_filter = [i.lower() for i in word.split()]
    tweets.append((word_filter, sentiment))
```

```
    #Calls above functions - gives us list of the words in the tweets, ordered by freq.
    #print getwordfeatures(getwords(tweets))

wordlist = getwordfeatures(getwords(tweets))
    wordlist = [i for i in wordlist if not i in stopwords.words('english')]
    wordlist = [i for i in wordlist if not i in customstopwords]

    #Creates a training set - classifier learns distribution of true/falses in the input.
    training_set = nltk.classify.apply_features(feature_extractor, tweets)
    classifier = nltk.NaiveBayesClassifier.train(training_set)

    #print classifier.show_most_informative_features(n=30)
    i = 0
    for line in open(sys.argv[1]).readlines():
        tweet = json.loads(line)
        tweet_score = 0
        if 'text' in tweet and 'lang' in tweet and tweet['lang'] == 'en':
            spectweet = tweet['text'].encode('utf-8')
            i = i + 1
            print 'Tweet No.',
            print i,
            print ':',
            spectweet = spectweet.replace('\n', '.')
            spectweet = spectweet.replace('\r', '.')
```

```
print spectweet
print 'Sentiment',
print i,
print ':',
print 'We think that the sentiment was ' +
classifier.classify(feature_extractor(spectweet)) + ' in that sentence.'
```

run.py

```
# python run.py > output.txt

import os
os.system('python NaiveBayes_LessTrainingTweets.py tweets.txt')
```

Appendix C: Support Vector Machines Approach Codes

SVM.py

```
# python SVM.py > output.txt; sed -i '1d' output.txt;

    import svm
    from svmutil import *
    import re, pickle, csv, os
    import json

    def getFeatureVector(tweet, stopWords):
        featureVector = []
        words = tweet.split()
        for w in words:
#replace two or more with two occurrences
w = replaceTwoOrMore(w)
            #strip punctuation
            w = w.strip('\'"?,.')
#check if it consists of only words
            val = re.search(r"^[a-zA-Z][a-zA-Z0-9]*[a-zA-Z]+[a-zA-Z0-9]*$", w)
#ignore if it is a stopWord
            if(w in stopWords or val is None):
                continue
```

```
            else:
                featureVector.append(w.lower())
        return featureVector

    def replaceTwoOrMore(s):
#look for 2 or more repetitions of character
        pattern = re.compile(r"(.)\1{1,}", re.DOTALL)
        return pattern.sub(r"\1\1", s)

    def processTweet(tweet):
        #Convert to lower case
tweet = tweet.lower()
#Convert www.* or https?://* to URL
tweet = re.sub('((www\.[^\s]+)|(https?://[^\s]+))','URL',tweet)
#Convert @username to AT_USER
        tweet = re.sub('@[^\s]+','AT_USER',tweet)
#Remove additional white spaces
tweet = re.sub('[\s]+', ' ', tweet)
#Replace #word with word
tweet = re.sub(r'#([^\s]+)', r'\1', tweet)
#trim
        tweet = tweet.strip('\'"')
        return tweet

    def getStopWordList(stopWordListFileName):
        #read the stopwords
```

```
    stopWords = []
    stopWords.append('AT_USER')
    stopWords.append('URL')

    fp = open(stopWordListFileName, 'r')
    line = fp.readline()
    while line:
        word = line.strip()
        stopWords.append(word)
        line = fp.readline()
    fp.close()
    return stopWords

featureList = [line.strip() for line in open('feature_list.txt')]

#Read the training tweets one by one and process it
inpTweets = csv.reader(open('training_dataset.csv'), delimiter=',', quotechar='|')
tweets = []
for row in inpTweets:
    sentiment = row[0]
    tweet = row[1]
    stopWords = getStopWordList('stopwords.txt')
    processedTweet = processTweet(tweet)
    featureVector = getFeatureVector(processedTweet, stopWords)
    tweets.append((featureVector, sentiment));
```

```
#Read the testing tweets one by one and process it
opTweets = csv.reader(open('testing_dataset.csv'), delimiter=',', quotechar='|')
test_tweets = []
for row in opTweets:
    sentiment = row[0]
    tweet = row[1]
    stopWords = getStopWordList('stopwords.txt')
    processedTweet = processTweet(tweet)
    featureVector = getFeatureVector(processedTweet, stopWords)
    test_tweets.append((featureVector));

def getSVMFeatureVectorandLabels(tweets, featureList):
    sortedFeatures = sorted(featureList)
    map = {}
    feature_vector = []
    labels = []
    for t in tweets:
        label = 0
        map = {}
        #Initialize empty map
        for w in sortedFeatures:
            map[w] = 0

        tweet_words = t[0]
        tweet_opinion = t[1]
        #Fill the map
```

```
for word in tweet_words:
            #process the word (remove repetitions and punctuations)
word = replaceTwoOrMore(word)
            word = word.strip('\'"?,.')
#set map[word] to 1 if word exists
            if word in map:
                map[word] = 1
        values = map.values()
        feature_vector.append(values)
        if(tweet_opinion == 'positive'):
            label = 0
        elif(tweet_opinion == 'negative'):
            label = 1
        elif(tweet_opinion == 'neutral'):
            label = 2
        labels.append(label)
    #return the list of feature_vector and labels
    return {'feature_vector' : feature_vector, 'labels': labels}


def getSVMFeatureVector(test_tweets, featureList):
        sortedFeatures = sorted(featureList)
        map = {}
        feature_vector = []
        for t in test_tweets:
            label = 0
            map = {}
```

```
#Initialize empty map
                for w in sortedFeatures:
                    map[w] = 0
                 #Fill the map
                for word in t:
                    if word in map:
                        map[word] = 1
                values = map.values()
                feature_vector.append(values)
            return feature_vector

    #Train the classifier
    result = getSVMFeatureVectorandLabels(tweets, featureList)
    problem = svm_problem(result['labels'], result['feature_vector'])
param = svm_parameter('-q')
    param.kernel_type = LINEAR
    classifier = svm_train(problem, param)



#Test the classifier
test_feature_vector = getSVMFeatureVector(test_tweets, featureList)
#p_labels contains the final labeling result
    p_labels, p_accs, p_vals = svm_predict([0] *
len(test_feature_vector),test_feature_vector, classifier)
```

```
count = 0
testing_tweets = []
for line in open('tweets.txt').readlines():
        tweet = json.loads(line)
        if 'text' in tweet and 'lang' in tweet and tweet['lang'] == 'en':
                spectweet = tweet['text'].encode('utf-8')
                spectweet = spectweet.replace('\n', '.')
                spectweet = spectweet.replace('\r', '.')
                testing_tweets.append(spectweet)


for i in p_labels:
    count = count + 1
    print 'Tweet No.',
    print count,
    print ':',
    print testing_tweets[count-1]
    print 'Sentiment',
    print count,
    print ':',
    if int(i) == 0:
        print 'We think that the sentiment was positive in that sentence.'
    if int(i) == 1:
        print 'We think that the sentiment was negative in that sentence.'
    if int(i) == 2:
        print 'We think that the sentiment was neutral in that sentence.'
```

run.py

 # python run.py

```
import os

os.system('python SVM.py > output.txt; sed -i \'1d\' output.txt;')
```

References

[1] S. Chinthala, R. Mande, S. Manne, and S. Vemuri, \Sentiment analysis on twitter streaming data," in Emerging ICT for Bridging the Future - Proceedings of the 49th Annual Convention of the Computer Society of India (CSI) Volume 1 (S. C. Satapathy, A. Govardhan, K. S. Raju, and J. K. Mandal, eds.), vol. 337 of Advances in Intelligent Systems and Computing, pp. 161{168, Springer International Publishing, 2015.

[2] G. Paltoglou, \Sentiment analysis in social media," in Online Collective Action (N. Agarwal, M. Lim, and R. T. Wigand, eds.), Lecture Notes in Social Networks, pp. 3{17, Springer Vienna, 2014.

[3] J. Dickerson, V. Kagan, and V. Subrahmanian, \Using sentiment to detect bots on twitter: Are humans more opinionated than bots?," in Advances in Social Networks Analysis and Mining (ASONAM), 2014 IEEE/ACM International Conference on, pp. 620{627, Aug 2014.

[4] K. Singhal, B. Agrawal, and N. Mittal, \Modeling indian general elections: Sentiment analysis of political twitter data," in Information Systems Design and Intelligent Applications (J. K. Mandal, S. C. Satapathy, M. Kumar Sanyal, P. P. Sarkar, and A. Mukhopadhyay, eds.), vol. 339 of Advances in Intelligent Systems and Computing, pp. 469{477, Springer India, 2015.

[5] \Python Documentation." https://www.python.org/doc/.
[Online; Last accessed on 07-April-2015].

[6] \Natural Language Toolkit, NLTK 3.0."
http://www.nltk.org/.
[Online; Last accessed on 06-April-2015].

[7] \LIBSVM." http://www.csie.ntu.edu.tw/~cjlin/libsvm/.
[Online; Last accessed on 06-April-2015].

[8] \Sentiment Dictionary." http://www2.imm.dtu.dk/pubdb/views/publication_details.php?id=6010.
[Online; Last accessed on 06-April-2015].

[9] \Naive Bayes Classi_er."
http://en.wikipedia.org/wiki/Naive_Bayes_classifier.
[Online; Last accessed on 06-April-2015].

[10] \Naive Bayes Classi_er." http://sentiment.
christopherpotts.net/classifiers.html#nb.
[Online; Last accessed on 06-April-2015].

[11] \Training Data Set." http://goo.gl/DUkjbQ.
[Online; Last accessed on 07-April-2015].

[12] \Support Vector Machines."
http://en.wikipedia.org/wiki/Support_vector_machine.
[Online; Last accessed on 07-April-2015].

[13] N. F. da Silva, E. R. Hruschka, and E. R. H. Jr., \Tweet
sentiment analysis with classi_er ensembles," Decision Support
Systems, vol. 66, no. 0, pp. 170 { 179, 2014.

[14] M. Erdmann, K. Ikeda, H. Ishizaki, G. Hattori, and
Y. Takishima, \Feature based sentiment analysis of tweets in
multiple languages," in Web Information Systems Engineering {
WISE 2014 (B. Benatallah, A. Bestavros, Y. Manolopoulos,
A. Vakali, and Y. Zhang, eds.), vol. 8787 of Lecture Notes in
Computer Science, pp. 109{124, Springer International
Publishing, 2014.

[15] H. Saif, Y. He, and H. Alani, \Semantic sentiment analysis of
twitter," in The Semantic Web { ISWC 2012 (P. Cudr_e-Mauroux,
J. Hein, E. Sirin, T. Tudorache, J. Euzenat, M. Hauswirth,
J. Parreira, J. Hendler, G. Schreiber, A. Bernstein, and
E. Blomqvist, eds.), vol. 7649 of Lecture Notes in Computer
Science, pp. 508{524, Springer Berlin Heidelberg, 2012.

[16] M. Ghiassi, J. Skinner, and D. Zimbra, \Twitter brand
sentiment analysis: A hybrid system using n-gram analysis and
dynamic arti_cial neural network," Expert Systems with
Applications, vol. 40, no. 16, pp. 6266 { 6282, 2013.